

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/873,224DATE: 01/11/2002  
TIME: 22:32:18

INPUT SET: S36717.raw

This Raw Listing contains the General  
Information Section and up to the first 5 pages.

## SEQUENCE LISTING

## (1) General Information:

## (i) APPLICANT:

(A) NAME: Innogenetics sa.  
(B) STREET: Industriepark Zwijnaarde 7, box 4  
(C) CITY: Ghent  
(E) COUNTRY: Belgium  
(F) POSTAL CODE (ZIP): B-9052  
(G) TELEPHONE: 00 32 9 241 07 11  
(H) TELEFAX: 00 32 9 241 07 99

ENTERED

(ii) TITLE OF INVENTION: New sequences of hepatitis C virus  
genotypes for diagnosis, prophylaxis and therapy.

(iii) NUMBER OF SEQUENCES: 270

## (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 09/873,224  
(B) FILING DATE:  
(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/362,455  
(B) FILING DATE:

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 213 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/873,224

DATE: 01/11/2002  
TIME: 22:32:18

INPUT SET: S36717.raw

```

47      (iii) ANTI-SENSE: NO
48
49      (vii) IMMEDIATE SOURCE:
50            (B) CLONE: BR34-4-20
51
52      (ix) FEATURE:
53            (A) NAME/KEY: CDS
54            (B) LOCATION: 1..213
55
56      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
57
58      CTC ACG GAA CGG CTT TAC TGC GGG GGC CCT ATG TTC AAC AGC AAG GGG      48
59      Leu Thr Glu Arg Leu Tyr Cys Gly Gly Pro Met Phe Asn Ser Lys Gly
60            1              5              10              15
61
62      GCC CAG TGT GGT TAT CGC CGC TGC CGT GCC AGT GGA GTT CTG CCT ACC      96
63      Ala Gln Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Pro Thr
64            20              25              30
65
66      AGC TTC GGC AAC ACA ATC ACT TGC TAC ATC AAG GCC ACA GCG GCT GCA      144
67      Ser Phe Gly Asn Thr Ile Thr Cys Tyr Ile Lys Ala Thr Ala Ala Ala
68            35              40              45
69
70      AGG GCC GCA GGC CTC CGG AAC CCG GAC TTT CTT GTC TGC GGA GAT GAT      192
71      Arg Ala Ala Gly Leu Arg Asn Pro Asp Phe Leu Val Cys Gly Asp Asp
72            50              55              60
73
74      CTG GTC GTG GTG GCT GAG AGT      213
75
76
77
78      Leu Val Val Val Ala Glu Ser
79            65              70
80
81      (2) INFORMATION FOR SEQ ID NO: 2:
82
83            (i) SEQUENCE CHARACTERISTICS:
84                  (A) LENGTH: 71 amino acids
85                  (B) TYPE: amino acid
86                  (D) TOPOLOGY: linear
87
88            (ii) MOLECULE TYPE: protein
89
90            (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
91
92      Leu Thr Glu Arg Leu Tyr Cys Gly Gly Pro Met Phe Asn Ser Lys Gly
93            1              5              10              15
94
95      Ala Gln Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Pro Thr
96            20              25              30
97
98      Ser Phe Gly Asn Thr Ile Thr Cys Tyr Ile Lys Ala Thr Ala Ala Ala
99            35              40              45

```

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/873,224

DATE: 01/11/2002  
TIME: 22:32:19

INPUT SET: S36717.raw

100

101 Arg Ala Ala Gly Leu Arg Asn Pro Asp Phe Leu Val Cys Gly Asp Asp  
102 50 55 60

103

104 Leu Val Val Val Ala Glu Ser  
105 65 70

106

107 (2) INFORMATION FOR SEQ ID NO: 3:

108

109 (i) SEQUENCE CHARACTERISTICS:

110 (A) LENGTH: 213 base pairs

111 (B) TYPE: nucleic acid

112 (C) STRANDEDNESS: single

113 (D) TOPOLOGY: linear

114

115 (ii) MOLECULE TYPE: cDNA

116

117 (vii) IMMEDIATE SOURCE:

118 (B) CLONE: BR36-23-18

119

120 (ix) FEATURE:

121 (A) NAME/KEY: CDS

122 (B) LOCATION: 1..213

123

124 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

125

126 CTC ACG GAA CGG CTT TAC TGC GGG GGC CCT ATG TTC AAC AGC AAG GGG 48  
127 Leu Thr Glu Arg Leu Tyr Cys Gly Gly Pro Met Phe Asn Ser Lys Gly  
128 1 5 10 15

129

130 GCC CAG TGT GGT TAT CGC CGC TGC CGT GCC AGT GGA GTT CTG CCT ACC 96  
131 Ala Gln Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Pro Thr  
132 20 25 30

133

134 AGC TTC GGC AAC ACA ATC ACT TGC TAC ATC AAG GCC ACA GCG GCT GCA 144  
135 Ser Phe Gly Asn Thr Ile Thr Cys Tyr Ile Lys Ala Thr Ala Ala Ala  
136 35 40 45

137

138 AGG GCC GCA GGC CTC CGG AAC CCG GAC TTT CTT GTC TGC GGA GAT GAT 192  
139 Arg Ala Ala Gly Leu Arg Asn Pro Asp Phe Leu Val Cys Gly Asp Asp  
140 50 55 60

141

142 CTG GTC GTG GTG GCT GAG AGT  
143 Leu Val Val Val Ala Glu Ser 213

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/873,224

DATE: 01/11/2002  
TIME: 22:32:19

INPUT SET: S36717.raw

```

153          (B) TYPE: amino acid
154          (D) TOPOLOGY: linear
155
156          (ii) MOLECULE TYPE: protein
157
158          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
159
160      Leu Thr Glu Arg Leu Tyr Cys Gly Gly Pro Met Phe Asn Ser Lys Gly
161          1              5              10              15
162
163      Ala Gln Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Pro Thr
164          20              25              30
165
166      Ser Phe Gly Asn Thr Ile Thr Cys Tyr Ile Lys Ala Thr Ala Ala Ala
167          35              40              45
168
169      Arg Ala Ala Gly Leu Arg Asn Pro Asp Phe Leu Val Cys Gly Asp Asp
170          50              55              60
171
172      Leu Val Val Val Ala Glu Ser
173          65              70
174
175      (2) INFORMATION FOR SEQ ID NO: 5:
176
177          (i) SEQUENCE CHARACTERISTICS:
178              (A) LENGTH: 213 base pairs
179              (B) TYPE: nucleic acid
180              (C) STRANDEDNESS: single
181              (D) TOPOLOGY: linear
182
183          (ii) MOLECULE TYPE: cDNA
184
185          (iii) HYPOTHETICAL: NO
186
187          (iii) ANTI-SENSE: NO
188
189          (vii) IMMEDIATE SOURCE:
190              (B) CLONE: BR36-23-18
191
192          (ix) FEATURE:
193              (A) NAME/KEY: CDS
194              (B) LOCATION: 1..213
195
196          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
197
198      CTC ACG GAG CGG CTT TAC TGC GGG GGC CCT ATG TTT AAC AGC AAG GGG      48
199      Leu Thr Glu Arg Leu Tyr Cys Gly Gly Pro Met Phe Asn Ser Lys Gly
200          1              5              10              15
201
202      GCC CAG TGT GGT TAT CGC CGT TGC CGT GCC AGT GGA GTT CTG CCT ACC      96
203      Ala Gln Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Pro Thr
204          20              25              30
205

```

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/873,224

DATE: 01/11/2002  
TIME: 22:32:20

INPUT SET: S36717.raw

```

206 AGC TTC GGC AAC ACA ATC ACT TGT TAC ATC AAA GCC ACA GCG GCC GCA      144
207 Ser Phe Gly Asn Thr Ile Thr Cys Tyr Ile Lys Ala Thr Ala Ala Ala
208           35                      40                      45
209
210 AAA GCC GCA GGC CTC CGG AGC CCG GAC TTT CTT GTC TGC GGA GAT GAT      192
211 Lys Ala Ala Gly Leu Arg Ser Pro Asp Phe Leu Val Cys Gly Asp Asp
212           50                      55                      60
213
214
215
216 CTG GTC GTG GTG GCT GAG AGT      213
217 Leu Val Val Val Ala Glu Ser
218           65                      70
219
220
221 (2) INFORMATION FOR SEQ ID NO: 6:
222
223     (i) SEQUENCE CHARACTERISTICS:
224         (A) LENGTH: 71 amino acids
225         (B) TYPE: amino acid
226         (D) TOPOLOGY: linear
227
228     (ii) MOLECULE TYPE: protein
229
230     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
231
232 Leu Thr Glu Arg Leu Tyr Cys Gly Gly Pro Met Phe Asn Ser Lys Gly
233   1                      5                      10                      15
234
235 Ala Gln Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Pro Thr
236           20                      25                      30
237
238 Ser Phe Gly Asn Thr Ile Thr Cys Tyr Ile Lys Ala Thr Ala Ala Ala
239           35                      40                      45
240
241 Lys Ala Ala Gly Leu Arg Ser Pro Asp Phe Leu Val Cys Gly Asp Asp
242           50                      55                      60
243
244 Leu Val Val Val Ala Glu Ser
245           65                      70
246
247 (2) INFORMATION FOR SEQ ID NO: 7:
248
249     (i) SEQUENCE CHARACTERISTICS:
250         (A) LENGTH: 213 base pairs
251         (B) TYPE: nucleic acid
252         (C) STRANDEDNESS: single
253         (D) TOPOLOGY: linear
254
255     (ii) MOLECULE TYPE: cDNA
256
257     (iii) HYPOTHETICAL: NO
258

```

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
**PATENT APPLICATION US/09/873,224**

DATE: 01/11/2002  
TIME: 22:32:20

**INPUT SET: S36717.raw**

Line	Error	Original Text
5	Mandatory Value Not Present	(i) APPLICANT:
6	Unknown or Misplaced Identifier	(A) NAME: Innogenetics sa.
7	Unknown or Misplaced Identifier	(B) STREET: Industriepark Zwijnaarde 7, box 4
8	Unknown or Misplaced Identifier	(C) CITY: Ghent
9	Unknown or Misplaced Identifier	(E) COUNTRY: Belgium
10	Unknown or Misplaced Identifier	(F) POSTAL CODE (ZIP): B-9052
11	Unknown or Misplaced Identifier	(G) TELEPHONE: 00 32 9 241 07 11
12	Unknown or Misplaced Identifier	(H) TELEFAX: 00 32 9 241 07 99

PAGE: 1

**SEQUENCE MISSING ITEM REPORT**  
**PATENT APPLICATION US/09/873,224**

DATE: 01/11/2002  
TIME: 22:32:20

*INPUT SET: S36717.raw*

ADDRESSEE  
STREET  
CITY  
STATE  
COUNTRY  
ZIP  
CORRESPONDENCE ADDRESS

SEQUENCE CORRECTION REPORT  
PATENT APPLICATION US/09/873,224DATE: 01/11/2002  
TIME: 22:32:20

INPUT SET: S36717.raw

Line	Original Text	Corrected Text
4037	(C) INDIVIDUAL ISOLATE: BR33 and BR36	(C) INDIVIDUAL ISOLATE: BR33 and BR36
4061	(C) INDIVIDUAL ISOLATE: HD10	(C) INDIVIDUAL ISOLATE: HD10
4086	(C) INDIVIDUAL ISOLATE: BR36	(C) INDIVIDUAL ISOLATE: BR36
4110	(C) INDIVIDUAL ISOLATE: BR36	(C) INDIVIDUAL ISOLATE: BR36
4137	(C) INDIVIDUAL ISOLATE: HD10	(C) INDIVIDUAL ISOLATE: HD10
4188	(C) INDIVIDUAL ISOLATE: BR36	(C) INDIVIDUAL ISOLATE: BR36
4215	(C) INDIVIDUAL ISOLATE: HD10	(C) INDIVIDUAL ISOLATE: HD10
5305	(C) INDIVIDUAL ISOLATE: GB358	(C) INDIVIDUAL ISOLATE: GB358
5327	(C) INDIVIDUAL ISOLATE: GB549	(C) INDIVIDUAL ISOLATE: GB549
5349	(C) INDIVIDUAL ISOLATE: GB809	(C) INDIVIDUAL ISOLATE: GB809
5370	(C) INDIVIDUAL ISOLATE: GB358	(C) INDIVIDUAL ISOLATE: GB358
5393	(C) INDIVIDUAL ISOLATE: GB549	(C) INDIVIDUAL ISOLATE: GB549
5414	(C) INDIVIDUAL ISOLATE: GB809	(C) INDIVIDUAL ISOLATE: GB809
5436	(C) INDIVIDUAL ISOLATE: GB358	(C) INDIVIDUAL ISOLATE: GB358
5459	(C) INDIVIDUAL ISOLATE: GB549	(C) INDIVIDUAL ISOLATE: GB549
5481	(C) INDIVIDUAL ISOLATE: GB809	(C) INDIVIDUAL ISOLATE: GB809
5503	(C) INDIVIDUAL ISOLATE: GB358	(C) INDIVIDUAL ISOLATE: GB358
5524	(C) INDIVIDUAL ISOLATE: GB549	(C) INDIVIDUAL ISOLATE: GB549
5545	(C) INDIVIDUAL ISOLATE: GB809	(C) INDIVIDUAL ISOLATE: GB809
5567	(C) INDIVIDUAL ISOLATE: GB358 and GB809	(C) INDIVIDUAL ISOLATE: GB358 and GB809
5590	(C) INDIVIDUAL ISOLATE: GB549	(C) INDIVIDUAL ISOLATE: GB549
5612	(C) INDIVIDUAL ISOLATE: GB549	(C) INDIVIDUAL ISOLATE: GB549